

VISIT...

LANZAROTE
Caliente.COM



Listen, you are writing! Speeding up online spelling with a dynamic auditory BCI

Martijn Schreuder, Thomas Rost and Michael Tangermann

Journal Name: Frontiers in Neuroscience

ISSN: 1662-453X

Article type: Original Research Article

Received on: 20 May 2011

Accepted on: 01 Sep 2011

Provisional PDF published on: 01 Sep 2011

Frontiers website link: www.frontiersin.org

Citation: Schreuder M, Rost T and Tangermann M(2011) Listen, you are writing! Speeding up online spelling with a dynamic auditory BCI. *Front. Neurosci.* 5:112. doi:10.3389/fnins.2011.00112

Article URL: http://www.frontiersin.org/Journal/Abstract.aspx?s=763&name=neuroprosthetics&ART_DOI=10.3389/fnins.2011.00112

(If clicking on the link doesn't work, try copying and pasting it into your browser.)

Copyright statement: ? 2011 Schreuder, Rost and Tangermann. This is an open-access article subject to a non-exclusive license between the authors and Frontiers Media SA, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and other Frontiers conditions are complied with.

This Provisional PDF corresponds to the article as it appeared upon acceptance, after rigorous peer-review. Fully formatted PDF and full text (HTML) versions will be made available soon.

Listen, you are writing!

Speeding up online spelling with a dynamic auditory BCI

Martijn Schreuder^{*,a}, Thomas Rost^{a,b}, Michael Tangermann^a

^a*Machine Learning Laboratory, Berlin Institute of Technology, Berlin, Germany*

^b*Bernstein Center for Computational Neuroscience, Berlin, Germany*

Abstract

Representing an intuitive spelling interface for Brain-Computer Interfaces (BCI) in the auditory domain is not straightforward. In consequence, all existing approaches based on event-related potentials (ERP) rely at least partially on a visual representation of the interface. This online study introduces an auditory spelling interface that eliminates the necessity for such a visualization. In up to two sessions, a group of healthy subjects (N=21) was asked to use a text entry application, utilizing the spatial cues of the AMUSE paradigm (**A**uditory **M**ulticlass **S**patial **E**RP). The speller relies on the auditory sense both for stimulation and the core feedback. Without prior BCI experience, 76% of the participants were able to write a full sentence during the first session. By exploiting the advantages of a newly introduced dynamic stopping method, a maximum writing speed of 1.41 characters/minute (7.55 bits/minute) could be reached during the second session (average: .94 char/min, 5.26 bits/min). For the first time, the presented work shows that an auditory BCI can reach performances similar to state-of-the-art visual BCIs based on covert attention. These results represent an important step towards a purely auditory BCI.

Key words: Brain-Computer Interface (BCI), directional hearing, auditory event-related potentials (ERP), P300, N200, dynamic subtrials

Introduction

Recent successes have been booked in the application of brain-computer interface (BCI) technology at the end-user's home, both for communication (Sellers et al., 2010; Nijboer et al., 2008) and other purposes (Münssinger et al., 2010). BCIs allow for a direct connection between the brain and the external world. Interpretation of the brain signals is generally performed in realtime and users can thus interact with a device by changing their brain state. Apart from applications for healthy users (Tangermann et al., 2009; Nijholt et al., 2009; Blankertz et al., 2010), BCI technology was ultimately meant to allow otherwise paralyzed people to communicate and interact with

their environment again.

Most of the aforementioned studies are based on visual event-related potentials (ERP). Their use in BCI was introduced by Farwell and Donchin (1988) and they have been a major focus in BCI research. This study describes a BCI that is based on auditory ERPs, which may extend the work on visual ERPs in several important aspects.

In an exploratory study (Huggins et al., 2011), 61 people suffering from mid-stage amyotrophic lateral sclerosis (ALS) were included with ALSFRS-R ratings ranging from 18 to 33 (Cedarbaum et al., 1999). The authors found that a larger portion of their population suffered from auditory- (42%) than visual (25%) deficits. On the other hand, a recent study which followed a patient with ALS through the late stages of the disease, reported that a disease-resistant BCI may only be possible through the auditory or proprioceptive pathways (Murguialday et al.,

*corresponding author: schreuder@tu-berlin.de

2010). Indeed, several studies show that the traditional visual BCI paradigms may not function well when eye-gaze control is limited (Treder and Blankertz, 2010; Brunner et al., 2010). A user group with such heterogeneous abilities and needs clearly stipulates the necessity for a range of diverse BCI systems. For these reasons, the interest in additional BCI paradigms has grown in recent years.

Some studies report that by using a different interface, a visual ERP BCI can be realized that does not depend on eye gaze control (Acqualagna et al., 2010; Liu et al., 2011; Treder et al., 2011). Another approach could be to switch to a different sensory domain altogether, by using auditory- (Hinterberger et al., 2004; Nijboer et al., 2008; Kanoh et al., 2008; Vlek et al., 2011; Kim et al., 2011) or vibrotactile cues (Cincotti et al., 2007; Brouwer and van Erp, 2010) for feedback or stimulation. Auditory stimulation usually involves an oddball paradigm where the user is exposed to stimuli that differ from each other on some property. The user is required to focus attention to one of them. Such paradigms generally involve tones with differences in pitch, resulting in a binary BCI (Hill et al., 2005; Halder et al., 2010). Multi-class approaches have been described using different environmental sounds (Klobassa et al., 2009) or spoken words or numbers (Sellers and Donchin, 2006; Furdea et al., 2009; Kübler et al., 2009; Guo et al., 2010). Though this increases the number of options per selection, presentation of such stimuli is inherently slower. Furthermore, the mapping is often not intuitive, so that the matrix is still shown (albeit not flashed).

Of special interest is a recent study (Kübler et al., 2009), which describes a patient trial with the paradigm previously reported in (Furdea et al., 2009). Four subjects suffering from ALS received extensive training with the auditory BCI. Three of them were close to, or in the locked-in state (LIS, only residual eye-movement), whereas the fourth was entering the totally locked-in state (TLIS, no residual eye-movement). Though all four subjects had control over a visual BCI, they performed poorly on the auditory BCI. The authors acknowledge however the importance of the auditory modality for BCI. One of their suggestions for improvement is a spatial distribution of the stimuli.

Several studies indeed show that using stimuli with spatial features can improve performance (Schreuder et al., 2009; Höhne et al., 2010). In offline studies, adding spa-

tial information to an auditory stimulus was beneficial to the recognition, both expressed in reaction times or accuracy (Belitski et al., 2011) and classification performance (Schreuder et al., 2010). Furthermore, it allows for an intuitive multi-class auditory BCI without sacrificing speed. Such paradigms could even be used to extend traditional (visual) BCIs to improve performance or facilitate training during the late stages of ALS (Belitski et al., 2011).

The current study is an extension of Schreuder et al. (2010), where the principle of a paradigm using spatial auditory stimuli was first shown (it was later called **Auditory Multiclass Spatial ERP**, or AMUSE). Here, AMUSE drives a spelling interface in online mode. Using six spatial locations and a two-step, hex-o-spell like speller interface (Blankertz et al., 2006), subjects were enabled to write by attending tones in space.

It is shown here that AMUSE, combined with the speller interface, allows for reliable online spelling for a majority of healthy subjects. The proposed speller interface is intuitive and easy to learn. Its core components (stimulation, spelling tree navigation and result feedback) are realized in the auditory domain. Average spelling performance was high, out-performing current auditory BCIs. Furthermore, a new method for dynamically changing the number of stimuli was introduced. Based on the data gathered, a trial could thus either be stopped if a threshold was reached, or continued. This method further increased the performance significantly by reducing trial length.

Methods

Participants

Participants were 21 subjects that had never worked with a BCI before (BCI naïve). They reported no current or prior neurological disorder and normal hearing. The latter was not formally tested. Subjects were financially compensated for their participation. Age ranged from 20 to 57 (mean = 34.1, SD = 11.4). Procedures were approved by the Ethics Committee of the Charité University Hospital. All subjects provided verbal and written informed consent and subsequent analysis and presentation of data was anonymized.

Table 1: **Stimulus properties.** The stimulus of each direction (see Figure 1) was a complex of a tone (base frequency and harmonics) and noise.

Direction	Base tone [Hz]	Noise range [KHz]
1	762	3.6 - 9.2
2	528	3.2 - 8.0
3	1099	4.0 - 10.5
4	635	3.4 - 8.6
5	915	3.8 - 9.8
6	440	3.0 - 7.5

Data acquisition

EEG was recorded using a fixed set of 56 Ag/AgCl electrodes and BrainAmp amplifiers (Brain Products, Munich, Germany). Channels were referenced to the nose. Electrooculogram (EOG) was co-recorded with two bipolar channels. All impedances were kept below 15 k Ω . Only for subject VPfce did the impedance exceeded this threshold at the end of the session. The signals were sampled at 1 kHz and filtered by a hardware analog bandpass filter between 0.1 and 250 Hz before being digitized and stored for offline analyzes. For online use, the signal was low-pass filtered below 40 Hz, downsampled to 100 Hz and streamed to the online Berlin BCI system.

The stimulus presentation, the online Berlin BCI system and the offline analyzes were implemented in Matlab (Mathworks), making use of the Psychophysics Toolbox (Brainard, 1997) for multichannel audio presentation and the open-source text-to-speech system Mary (Schröder and Trouvain, 2003). A multichannel, low-latency firewire soundcard from M-Audio (M-Audio Firewire 410) was used to individually control the active, off-the-shelf computer speakers (type Sony SRS-A201).

Parameter screening

A dependence of ERP amplitude and latency on stimulus amplitude and duration, particularly in the early components, has previously been reported (Gonsalvez et al., 2007). To investigate this for the AMUSE paradigm, two parameter screenings were conducted, testing for the influence of stimulus loudness and duration on classification performance. The studies were performed with six and seven healthy subjects, respectively.

Results of both the parameter screening did not expose a systematic influence on the classification performance. Also, the subjectively reported *favourite* condition could not be matched with the condition giving the

highest score. Parameters were therefore fixed to those used in Schreuder et al. (2010, ~58 dB and 40 ms).

Stimuli

The localization of a stimulus in the azimuth plane depends mainly on two principles, interaural timing differences (ITD) and interaural level differences (ILD), which both have their own optimal frequency range (Middlebrooks and Green, 1991). Where the ITD effect is most efficient for frequencies up to 1-1.3 KHz, the ILD effect provides the localization information for higher frequencies, roughly >3 KHz. Offline analyzes showed that spatial location as a discriminative feature is enough to reliably elicit a P3 response (Schreuder et al., 2009, 2010). However, to optimally exploit both effects, stimuli consisted of a low frequent tone with harmonics and a high frequent, band-pass filtered noise overlay (see Table 1). To further increase resolution, each of the six directions was associated with a unique combination of tone and noise to add an additional cue property. Stimuli can be found in Data sheet 1 of the supplementary data.

Procedure

Subjects sat in a reclining chair, facing a screen with a fixation cross at ~1m. They were surrounded by six speakers at ear height, evenly distributed in a circle with 60° distance between them (see Figure 1-A) which is well over the resolution of spatial hearing in the azimuth plane (Grantham et al., 2003). Circle radius was ~65 cm and speakers were calibrated to a common stimulus intensity of ~58 dB. The room was neither electromagnetically shielded, nor were extensive sound attenuating precautions taken. Before the experiments, subjects were asked to minimize eye movements and muscle contractions during recording periods.

The experimental protocol is visualized in Figure 1-B. A *run*, or sentence, consisted of several *trials*. Each *trial*, or selection, consisted of several *iterations*. During one iteration all speaker locations were stimulated exactly once in a pseudo-random order, before proceeding to the next iteration. This pseudo-random sequence was generated such that (1) between two stimuli from one speaker, there were at least two other stimuli and, (2) two successive stimuli never came from neighboring speakers. A single stimulus is hereafter referred to as a *subtrial*. The stimulus onset asynchrony (SOA) was set to 175 ms.

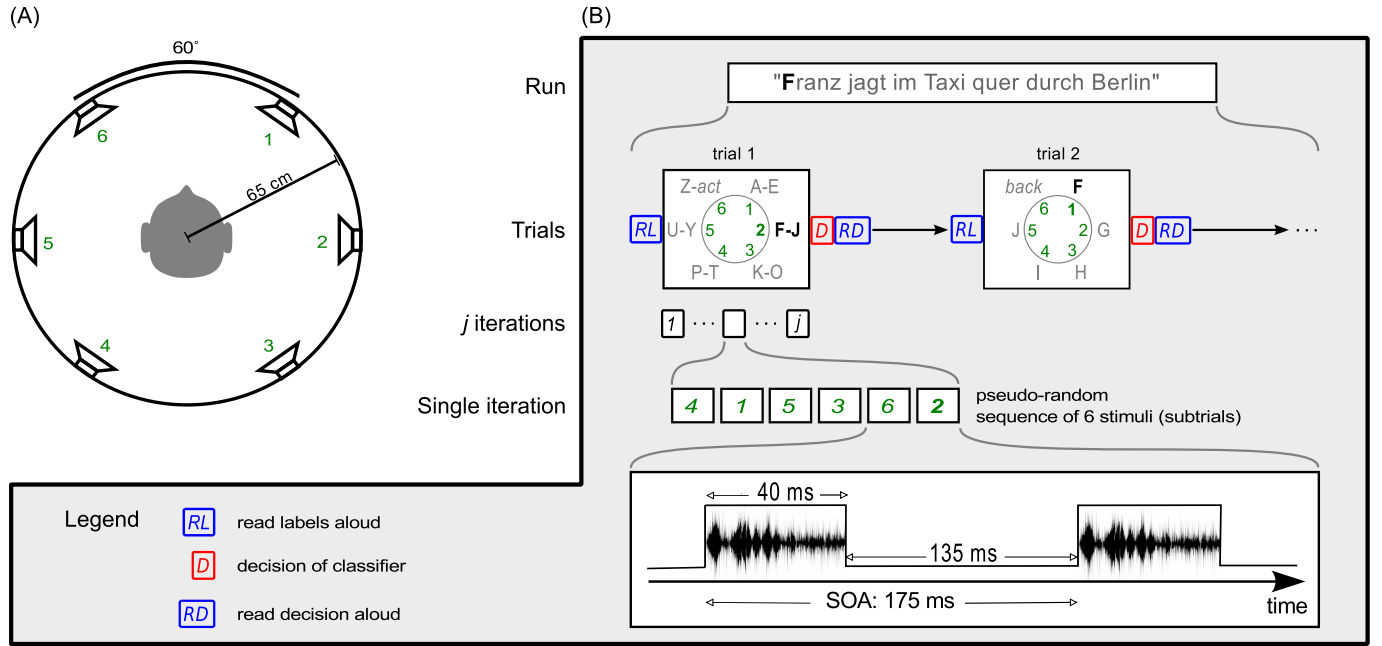


Figure 1: **Experimental protocol.** (A) Subjects were surrounded by six speakers at ear height. Speakers were equally spaced with 60° angle between neighbors, with a circle diameter of 65 cm. (B) Visualization of the experimental protocol in the online phase. In Session 1, the number of iterations j was set to 15. For Session 2, j was determined by the dynamic stopping method. No labels were read aloud (RL) in Session 2.

The amount of iterations/trial influences the efficiency of the system. A high number is likely to increase the accuracy, but at the same time it increases the time needed for a selection. Vice versa, a low number will allow for quicker selections, but they are likely to contain more errors. Therefore, two settings were tested: (1) a fixed number of 15 iterations and, (2) a method which allows for determining the number of iterations on runtime.

During a trial, the subject's task was to focus attention to one (target) of the six directions and mentally count the number of appearances. During the calibration phase the target was given to the subject explicitly. During the writing phase the target direction was not given explicitly, but had to be inferred from the spelling interface.

All 21 subjects performed a first session (Session 1), and those that reached a performance level that allowed them to write the given sentence were invited for a second session (Session 2).

Spelling interface

For online writing, an adapted version of the hex-spell speller (Blankertz et al., 2006) was created in which a character can be selected in two steps. First a group of characters is selected (group 'F-J' in Figure 1-B) by focusing on the corresponding direction. In the second step the characters of this group are divided over five of the directions and an individual letter can be selected. Choosing the sixth direction returns the user back to the first selection step, thus preventing a wrong letter selection. Which (group of) letter(s) corresponds to a direction – and which direction had been selected – could be read out to the user using speech synthesis (Schröder and Trouvain, 2003). The copy text and progress were presented visually.

Session 1

First, subjects were familiarized with the sounds, after which about 30 minutes of calibration data were recorded. Subjects performed 48 trials, eight for each direction. At

the start of a trial the target direction was indicated both visually (one of six on-screen dots highlighted) and auditory (by playing the direction specific cue from the target location). After that, stimulation was purely auditory and subjects were asked to count the number of target appearances in the trial and report them. Each calibration trial consisted of 15 iterations. However, in order to have a varying number of targets and thus have a more challenging task, a prequel consisting of varying length (1-3 iterations) was added. That the visual target presentation was not necessary was proven in a session with a single, blind subject.

Then, subjects were asked to write word for word one of two sentences: "FRANZ JAGT IM TAXI DURCH BERLIN" or "SYLVIA WAGT QUICK DEN JUX BEI PFORZHEIM". Sentence assignment was random. Wrong letters had to be corrected by the user by performing a backspace action, requiring two selections. No prequel was included. The stimulation sequence and the overhead of label-, and result presentation for a single trial took around 34 seconds. With a two step letter selection process, the maximum theoretical speed is thus .89 char/min.

Session 2

Subjects who were able to successfully write with the system in Session 1 came back for Session 2. This session consisted of a calibration phase – equal to Session 1 – and the online writing of two German sentences. For Session 2 the protocol was adjusted in several ways. First, a dynamic stopping method was introduced which could stop a trial after 4-15 iterations when enough evidence for a correct selection was found (see below). Second, subjects were asked to study the selections that are needed for each character at home. They were no longer provided with auditory labels prior to a trial, which reduced the trial time by about 6.2 seconds. The maximum theoretical speed with 15 iterations was thus 1.10 char/min.

During the first sentence, subjects got visual information on the labels to familiarize them again with the interface. During the second sentence this was turned off, thus relying exclusively on the auditory sense. During both sentences, subjects still received auditory information after a selection to know their current location in the spelling tree. The first sentence was the remaining sentence from Session 1, whereas the second sentence could

be chosen by the subject but should be at least five words long.

Features and classifier

Classification was done on each subtrial, thus reducing the problem to a binary task, i.e. to classify each subtrial as a target of non-target. Spatio-temporal features for classification were extracted according to the following procedure. First the data were baselined, using the 150 ms pre stimulus data as reference. Then, the sample-wise r^2 coefficient was calculated for targets vs non-targets. Based on this, a set of two to four intervals with high discriminative information content were hand-picked, such that both early and late components were represented. All samples within each interval were then averaged, so that one interval was represented by one feature per channel. This resulted in an overall dimensionality of 112 to 224 features.

The benefit of this method over conventional sub sampling (Krusienski et al., 2006; Schreuder et al., 2010) is a relatively small number of features per channel which capture most of the discriminative information. Furthermore, by combining consecutive samples instead of picking individual samples, as is the case in step-wise linear discriminant analysis (Farwell and Donchin, 1988), the interpretation of the chosen features is straight-forward and a-priori knowledge can be included during the selection. Using this feature vector, a linear, binary classifier was trained. In order to prevent over-fitting, the classifier was conditioned using shrinkage regularization (Ledoit and Wolf, 2004; Blankertz et al., 2011).

Decision making

The classifier was trained to assign negative scores to target subtrials and positive scores to non-target subtrials. Let $X \in \mathbb{R}^{C \times J}$ be a matrix of classifier scores of a trial, where C is the number of classes (six in our case), and J the number of performed iterations. If $c = \{1, \dots, C\}$, then let $\tilde{x}$ be a row vector, where $\tilde{x}_c$ denotes the median value of classifier scores for class c . The winning class c^* can be described as $c^* = \operatorname{argmin}_c \tilde{x}_c$, i.e. the class with the lowest median value. For Session 1 J was fixed to 15.

In order to further optimize the spelling speed, a dynamic stopping method was introduced in Session 2. Such a method allows for trials with less than 15 iterations,

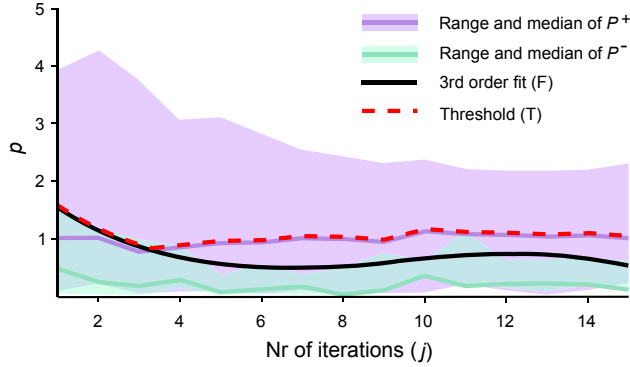


Figure 2: **Visualization of iteration-specific decision threshold.** The confidence measure p is plotted for correct (P^+) and incorrect (P^-) decisions as a function of the number of iterations j . The black line F is fitted to the maximum values of P^+ . The decision threshold T is defined as the maximum of F and the median of P^+ .

when the obtained data support an early stop. Different methods for tackling this have been introduced to BCI before (Serby et al., 2005; Lenhardt et al., 2008; Zhang et al., 2008; Liu et al., 2010), and were mainly tested in the context of visual ERP based BCI and one-step interfaces.

It can be assumed that at a low number of iterations the quality of a decision will be susceptible to sporadic outlying classifier scores, and that the influence of these will decrease with an increasing number of iterations. In line with this assumption, an approach which uses iteration specific thresholds was designed. Thresholds are based on the calibration data, and are thus subject specific.

The variable p is defined as the distance from $\tilde{x}_c$ to the second lowest value in $\tilde{x}$. It represents a confidence measure of the winning class. Note that, as this is a relative value, it is robust to linear biases that might arise when moving from the calibration phase to the online phase. In the online setting, the BCI checks after each iteration $j = 1 \dots J$ if p_j – based on the data gathered thus far – exceeds threshold T_j (see below). If so, the trial is stopped and class c is selected.

Heuristic for decision threshold determination. For finding the decision thresholds T , the trained classifier is re-applied to the complete calibration data. Although this can be considered overfitting and might lead to an over-estimation of the class distances, it is not considered as a caveat when finding a conservative threshold.

For calculating T_j , we consider all calibration data collected within a trial, up to iteration j . First, the winning class for this iteration c_j and the confidence p_j are calculated and p_j is assigned a label; true (p_j^+) if c_j corresponds to the target direction, and false (p_j^-) otherwise. This is repeated for each trial $n = 1 \dots N$ and iteration j , and results are collected in the matrices $P^+ \in \mathbb{R}^{N \times J}$ and $P^- \in \mathbb{R}^{N \times J}$. In order to approximate a smooth upper error bound, a third order polynomial F is fitted to the upper limit of P^+ . Then, if f_j refers to the value of F at iteration j , the iteration specific threshold T_j is defined as the maximum of f_j and the product of the median of column j in P^+ and R . The hyperparameter R is used to control the number of false negative stops. In this study, R was set to 1, which means that maximally only 50% of possible correct early stops is recognized. The above is visualized in Figure 2.

The threshold is rather conservative, as it is biased away from false positives (incorrect early stops). The rational behind such a conservative threshold is that any false positive is costly, as correcting an error takes up to four selections. For applications where a false selection is less costly, the threshold can be set less conservative by taking a smaller value for R .

Data Analyzes

The BCI performance evaluation was done online, unless stated otherwise. Three evaluation metrics were calculated for each subject and each sentence: accuracy (acc), characters per minute (char/min) and information transfer rate (ITR) (Schlögl et al., 2007). We define accuracy as the number of correct selections divided by the total amount of selections. Accuracy assesses the performance of the BCI, irrespective of the application interface.

For assessing the BCI in combination with the spelling application we used char/min as the preferred metric for writing proficiency. Char/min is defined as the number of correctly written letters divided by the time it took to write all letters (including incorrect letters). Here, it gives a particular conservative estimate, as subjects had to correct any mistakes. Thus, all letters are in the end correct at the cost of longer writing time (and lower char/min). This is a more realistic assessment of the system’s usefulness, as it includes the error recovery strategy. In a real user setting, errors will occur and may have to be corrected.

Table 2: **Performance summary.** Performance measures accuracy (*Acc*), characters per minute (*Char/min*) and information transfer rate (*ITR*) are given for all subjects and all sentences, as well as age and gender. Two means are calculated over each measure. *Session mean* refers to the mean when considering all available subjects, whereas *In all sessions* refers to the mean of only those subjects that succeeded in all sentences. The latter is included for cross session comparison. Subjects are sorted according to their accuracy on the first session. Missing values are either because a the sentence was not written by the user (x) or because the Char/min could not be calculated (-). The non-linear relation between accuracy and characters per minutes can be explained by the type of errors committed, i.e. on the first or second level.

User	Gender	Age	Session 1			Session 2					
			Sentence 1			Sentence 2a			Sentence 2b		
			Acc [%]	Char/min	ITR	Acc [%]	Char/min	ITR	Acc [%]	Char/min	ITR
VPfaz	F	27	98.4	.83	4.59	95.9	1.26	7.12	100.0	1.41	7.55
VPfcc	M	28	98.4	.84	4.64	85.1	.88	5.07	88.6	.88	5.38
VPkw	F	57	97.0	.81	4.51	86.5	.94	5.45	86.5	.95	5.68
VPfca	M	23	94.7	.77	4.35	97.0	1.30	7.32	94.2	1.14	6.66
VPfcd	M	25	92.9	.77	4.26	91.3	1.10	6.05	97.9	1.35	7.01
VPfaw	M	40	91.7	.73	4.15	92.7	1.01	5.96	94.3	1.03	6.05
VPfar	M	31	85.6	.58	3.55	89.8	.96	5.58	83.3	.76	4.70
VPfav	M	40	84.9	.59	3.51	88.3	1.01	5.94	96.8	1.38	7.27
VPfcb	M	21	84.7	.49	3.51	81.7	.65	4.45	85.1	.75	5.47
VPfck	M	43	83.2	.50	3.28	78.0	.61	4.20	68.5	.40	2.47
VPfau	F	25	81.9	.50	3.16	93.4	1.03	6.08	83.9	.74	4.74
VPfcj	F	23	80.6	.41	3.00	75.3	.50	3.68	69.3	.65	2.32
VPfcg	F	33	79.5	.40	3.20	77.8	.92	4.14	88.5	1.06	6.07
VPfch	M	55	79.3	.46	2.79	57.9	.36	0.64	x	x	x
VPfcm	M	51	78.3	.47	2.72	73.4	.51	2.72	68.1	.61	2.32
VPfax	M	28	72.1	.32	2.02	x	x	x	x	x	x
VPfci	M	28	58.6	-	1.47	x	x	x	x	x	x
VPfat	F	25	54.1	-	.58	x	x	x	x	x	x
VPfcl	F	44	51.7	-	.27	x	x	x	x	x	x
VPfce	M	47	40.0	-	.00	x	x	x	x	x	x
VPfas	F	29	37.5	-	.00	x	x	x	x	x	x
Session mean		34.1	77.4	.59	2.84	84.3	.87	4.96	86.1	.94	5.26
In all sessions mean		33.4	88.0	.62	3.75	86.2	.91	5.27	86.1	.94	5.26

For the sake of comparison, the ITR is also reported. We use the calculation method proposed by Schlögl et al. (2007), as there are systematic confusions between directions. ITR combines the accuracy, number of possible decisions and the number of selections per minute into a single number. Originating from information theory, it assesses information transfer over a noisy channel. Although often used in BCI, it is a theoretical measure and its interpretability in terms of practical relevance of a BCI is difficult.

Directional preference

The confusion matrix (CM) provides a lot of information about the discriminability of the different classes. From it, the pairwise F-scores were calculated to assess the confusion between each pair of directions. This should expose any systematic directional preference, if present.

Given a confusion matrix $CM_{N \times N}$, where N is the number of classes and each row is normalized, we define the

sensitivity and recall for pairs of classes $i, j \in \{1, \dots, 6\}$ as follows

$$(1) \quad sensitivity_{(i,j)} = \frac{CM_{(i,i)}}{CM_{(i,i)} + CM_{(i,j)} * (N - 1)}$$

$$(2) \quad recall_{(i,j)} = \frac{CM_{(i,i)}}{CM_{(i,i)} + CM_{(j,i)} * (N - 1)}$$

The matrix $F_{N \times N}$ of pairwise F-scores is then defined as

$$(3) \quad F_{(i,j)} = 2 \times \frac{sensitivity_{(i,j)} * recall_{(i,j)}}{sensitivity_{(i,j)} + recall_{(i,j)}}$$

where $1 \leq i \leq C$ and $1 \leq j \leq C$. It is thus the harmonic mean of sensitivity (the ability to identify a target) and recall (the ability to reject a non-target).

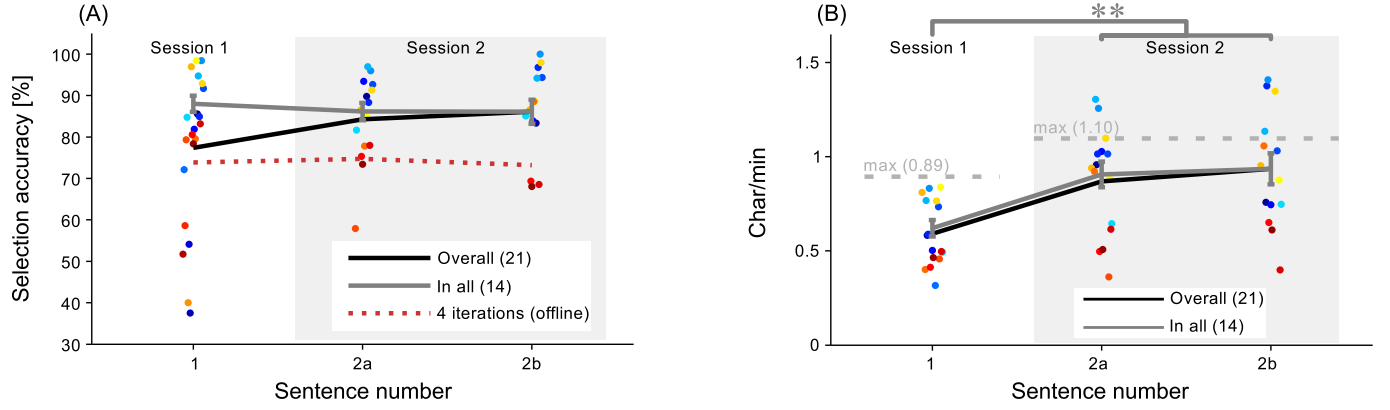


Figure 3: **Session differences.** Session differences in accuracy (A) and char/min (B) are plotted for all subjects. The black line represents the mean of all subjects in a session and the gray line represents the mean of the 14 subjects that succeeded on all three sentences. (A) No significant difference was found between sessions in the accuracy. No learning effect was found even when the number of iterations was reduced to four, as represented by the red, dotted line. (B) The char/min for Session 2 was significantly better than for Session 1, which is due to early stopping and less label vocalization. Some subjects performed higher than the theoretic maximum char/min, indicating a particular benefit from the dynamic stopping mechanism. Points are slightly displaced on the horizontal axis to prevent clutter.

Results

Of the 21 subjects, 16 subjects (or 76%) were able to write a full German sentence at first try (Session 1). The other five subjects got at most one word correct, before their experiment was stopped. In Session 2, 14 out of 15 returning subjects wrote both full German sentences. Subject VPfch was able to write the first sentence, but took so long that the session had to be aborted after this. Subject VPfax was unable to return for the second session, although successful in the first session. In the following, sentence 1 refers to the sentence written in the first session, and sentence 2a and 2b to those written in the second session. Individual performance measures for each subject can be found in Table 2. Movie 1 of the supplementary data shows an example of a user operating the system.

BCI performance

BCI performance did neither correlate with age ($r < 0.01$, $p = 0.75$), nor with gender (Kolmogorov-Smirnov test: $p = 0.42$). Average accuracy per sentence was 77.4%, 84.3% and 86.1% for sentence 1, 2a and 2b, respectively. The increase in performance from Session 1 to Session 2 is due to the drop out of less-performing subjects. This is shown in Figure 3-A, which gives the

overall accuracy for each subject and each session. Three means were calculated; first, the black line indicates the mean when considering all subjects for a session, as given above. However, as those subjects that were unable to write in the first session did not participate in the second session, the gray line includes only those subjects ($N=14$) that finished all three sentences, which is a fairer comparison of the sessions (.88, .86 and .86 for sentence 1,

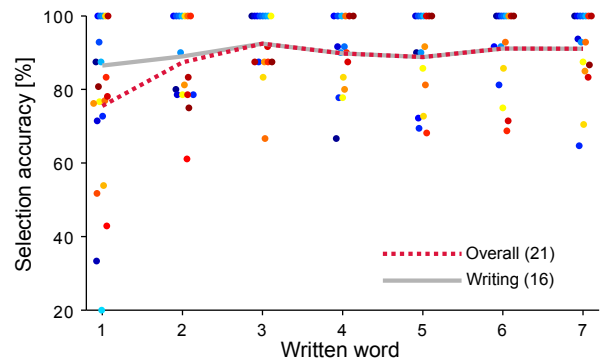


Figure 4: **Accuracy over time.** Selection accuracy of Session 1 is plotted for each subject as a function of the word written, which approximates time. The gray line represents the mean over the subjects that finished the entire sentence, which shows no significant slope. The red dotted line represents the mean over all subjects.

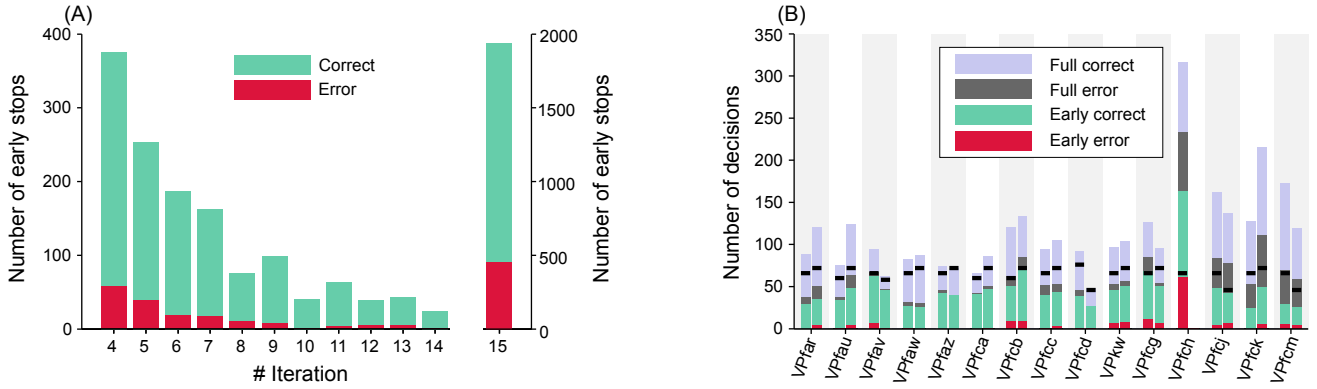


Figure 5: **Early stopping performance.** For (A), data is collapsed over all subjects that participated in Session 2. The number of correct and incorrect trials are plotted as a function of the iteration in which a trial was stopped. Accuracy is high in general, and particularly for those trials that were stopped before the 15th iteration. The same data is plotted in (B), but separately for each subject and for sentences 2a and 2b. The black markers indicate the minimum number of selections necessary for writing a sentence. Subject VPfch did not perform sentence 2b.

2a and 2b respectively). No significant difference existed ($p = .97$) between performance in sentence 2a ($m=.86$, $sd=.08$) and 2b ($m=.86$, $sd=.11$). The difference between these conditions is that for sentence 2b the visually presented support labels were switched off. This shows that the spelling tree can easily be learned and thus need not be presented visually.

As the number of iterations varied in Session 2, an offline analysis was performed to objectively assess any learning effect between session. For each sentence, the online classification scores of the first four iterations were used to take a decision (the minimum number of iterations in Session 2 was four). The red, dotted line in Figure 3-A shows these results. No systematic learning effect was present. As sentence 1 ($m=.88$, $sd=.07$) and sentence 2b showed no significant performance difference ($p = .40$), it can thus be concluded that early stopping did not introduce additional errors. Statistical testing was done using two-sided paired t-tests.

Figure 4 shows the accuracy for each subject as a function of the word written in the first sentence, which we take as an approximation of time. Considering only those subjects that were able to finish the session, the slope of the linear least-squares fit between performance and the word number is close to 0, implying that subjects are able to use the interface straight away.

The average ITR over all 21 subjects of Session 1 was

2.84 bits/min (maximum 4.59 bits/min), which includes also subjects that did not reach a performance level appropriate for writing (see Table 2). When considering only those subjects that performed well on all sentences, the average ITR over Session 1 was 3.75 bits/min. During the second session this increased by 40% to an average of 5.26 bits/min, with a maximum of 7.55 bits/min.

Writing efficiency

Figure 3-B presents the writing proficiency for each subject and each session in terms of characters per minute (char/min). The black line represents the sentence-wise mean for all subjects who finished that sentence. The gray line represents the sentence-wise mean for only those subjects that finished all sentences ($N=14$), which will be considered for further analyzes. As expected from the accuracy, there is no significant difference between sentence 2a ($m=.91$, $sd=.25$) and sentence 2b ($m=.94$, $sd=.31$) of Session 2 ($p = .57$). However, when comparing sentence 1 ($m=.62$, $sd=.16$) with sentence 2b, a significant increase of 52% can be observed ($t = 4.68$; $p < .001$). As the accuracy was equal, the increase must be due to improvements in the interface's efficiency. These were twofold: (1) dynamic stopping was introduced and (2) the vocalization of the labels was omitted. To assess the influence of the dynamic stopping method independently, the 6.2 seconds time gain from the second modification was added again

to sentence 2b in an online simulation. A significant increase of 37% remains from sentence 1 to the simulated sentence 2b ($m=.85$, $sd=.26$) ($t = 4.05$; $p = .001$). Thus, the dynamic stopping method increases performance significantly by itself. Statistical testing was done using two-sided paired t-tests.

Dynamic stopping method

Over all subjects, a total of 3297 online trials were recorded in Session 2. The dynamic method produced early stops in 41.3% of these trials. The maximum number of 15 iterations (full trial) was reached in the remaining 58.7% of the trials. This was in cases where the threshold was not exceeded in iterations 4 – 14. Early stops before the fourth iteration were not allowed.

About 12.6% of early stops resulted in an incorrect decisions, which is only 5.2% of the overall number of trials. This reflects the conservative threshold policy as described in the *methods*. At 23.5%, the percentage of errors was almost twice as high for full trials. An obvious explanation is that mostly the 'difficult-to-decide-trials' reach the full number of iterations.

As depicted in Figure 5-A, the majority of early stops is performed directly at the fourth iteration, with the number of stops decreasing until finally remaining trials are stopped at the 15th trial. At each individual iteration the percentage of incorrect decisions is lower than for full trials.

Figure 5-B gives the different errors for each subject and sentence 2a and 2b. It shows that the error rate for early stops is relatively low, and similar for all subjects. Only subject VPfch had such a high error rate – both for early stops and full trials – that there was no time left for writing the second sentence. The black lines indicate the minimum number of decisions necessary for writing the particular sentence.

Physiology

Calibration data from the first session of all subjects was pooled and the grand average calculated. Artifacts were rejected according to a fixed variance criterium and a peak-to-peak difference criterium. Figure 6-A shows the grand average time series for channel Cz. An early component is consistently present, peaking about 130 ms after each stimulus. This causes a rhythmic perturbation of

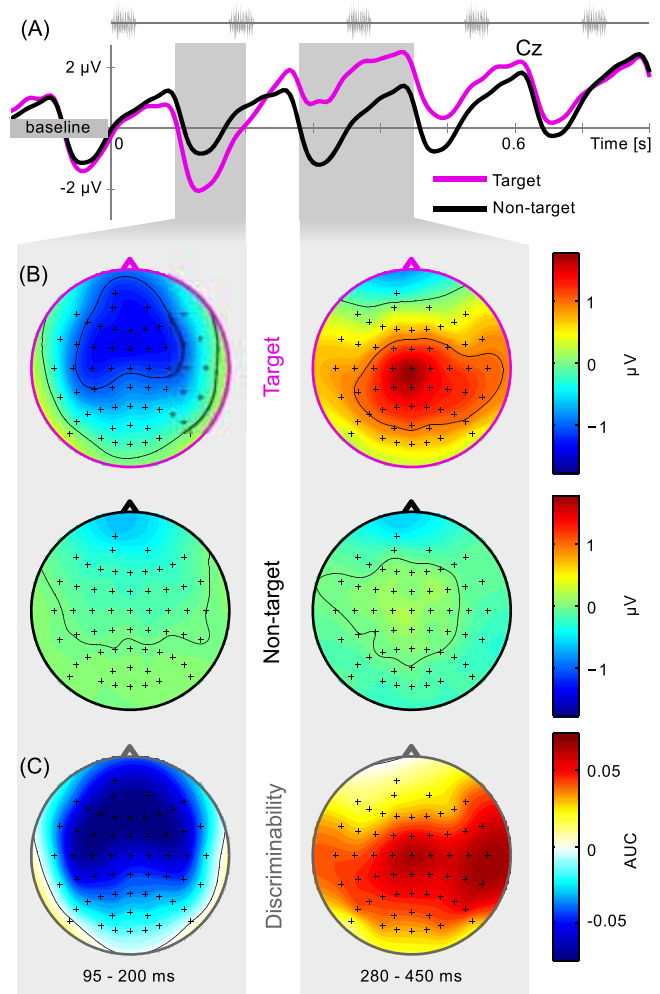


Figure 6: **Grand average physiology.** Calibration data of the first session is collapsed over subjects and trials. Average time series for channel Cz are plotted together with the positions of five consecutive stimuli (A). Scalp potential distributions for target and non-target trials are shown in (B) and distributions of class-discriminative information in (C). Two intervals were hand picked based on the signed AUC values, and are represented by the shaded gray areas. Early negative components over frontal and temporal regions are followed by a positive component over centro-parietal areas.

the EEG signal, thereby masking the generally described ERP components. Although present in both target and non-target trials, it is more negative for target trials, i.e. it is attention modulated. An interval from 95 to 200 ms

was selected to capture this peak.

A positive component is found for target trials, captured in the interval from 280 to 450 ms. Due to both timing and scalp location, we consider this to be what is described as the P3 component.

Figure 6-B shows the grand average scalp topography for both target and non-target trials for both selected intervals, whereas Figure 6-C displays the discriminability between those, expressed in signed AUC values (Schreuder et al., 2010). Low voltage amplitude and AUC values are due to inter subject differences. Online data are similar and thus not presented here. It is apparent that the early negative component is found mainly over the frontal and bilateral auditory cortical areas. The later positive component is focused over the central midline electrodes and toward the temporal areas. This is in line with previous offline results (Schreuder et al., 2010). It can be noted that in the manual feature extraction procedure, different intervals can be assessed based on this physiological plausibility before being included for online use.

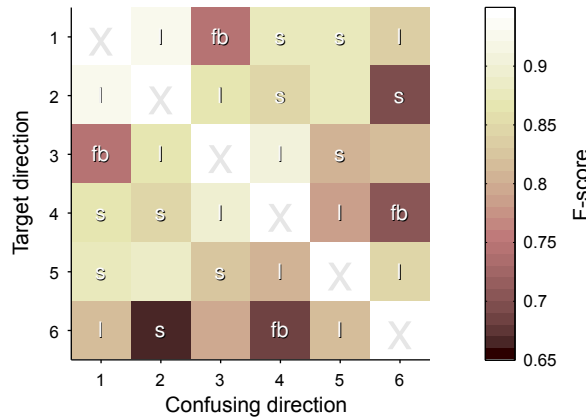


Figure 7: **Pairwise F-scores between speaker locations.** The matrix visualizes the pairwise confusions between an intended location (y axis) and a distracting location (x axis), expressed in F-scores. Lighter colors mean better discrimination. Entries are labeled as *l* = location neighbor, *s* = sound neighbor and *fb* = front-back symmetric location.

Directional preference

The grand-average confusion matrix (CM) can be found in the supplementary data (Table 3). Figure 7 displays matrix *F* of the pair-wise F-scores, where high values indicate a good separability between the indicated

classes. Diagonal values are irrelevant and marked X. Other elements are marked as location neighbors (*l*), sound neighbors with 1.5 tone pitch difference (*s*) and locations that are symmetric in the front-back plane (*fb*); all are expected to hamper separability. Tone locations were optimized such that no entry has two such labels.

Directions without any degrading factor had an average pairwise F-score of .89 (SD: .15). This was only slightly degraded for spatial neighbors (.87, SD: .16), but both pitch neighbors (.82, SD: .19) and especially the front-back confusion (.76, SD: .22) had a larger effect on pairwise discriminability. These effects were investigated by performing pairwise t-tests on the F-scores of each degrading factor with those pairs without degrading factors. After Bonferroni correction, the front-back confusion showed a trend of negative influence on recognition ($p = .056$).

From the pitch neighbors, the highest and lowest pair were most often confused. All this was in line with subjective reporting of the subjects, and gives indicators for future improvements.

Discussion

A previous offline study already showed the potential of the AMUSE paradigm (Schreuder et al., 2009, 2010). Now, coupled to a spelling interface, it allowed a majority of healthy subjects to write a full sentence with an average speed of .59 char/min on the first session, and up to a maximum speed of 1.41 char/min in the second session. To our knowledge, this is the fastest online auditory BCI speller described in literature. For the sake of comparison, it is important to emphasize once more the conservative nature of the protocol.

In the first session, 16 out of 21 subjects (or 76%) were able to write a full German sentence. According to Guger et al. (2009) over 90% of people is 'able to control a P300-based BCI'. However, as both their tested conditions were overt visual paradigms this may be a too optimistic estimate. The selected BCI paradigms are most prevalent in literature, but may not match fully with the user's needs and abilities (Murguialday et al., 2010; Treder and Blankertz, 2010; Brunner et al., 2010). Furthermore, for such overt ERP BCI paradigms the learning effect is generally small and average performance is high and consistent over subsequent trials. However, recently Klobassa

et al. (2009) showed that for their auditory P3 paradigm the performance increased significantly after 11 sessions. With subjects performing only up to two sessions, such an increase was not observed here. However, with more training the percentage of people able to use the setup could increase.

Performance

To assess the practical relevance of a BCI speller, the most intuitive metric is the number of characters that a user can write per minute. However, as this is not uniformly utilized in BCI literature, it makes direct comparison troublesome. For this reason, the ITR as defined by (Schlögl et al., 2007) was reported. The current setup without any dynamic stopping (Session 1) shows ITR values that are en par with the online results of the state-of-the-art auditory BCI systems (Sellers and Donchin, 2006; Furdea et al., 2009; Klobassa et al., 2009; Höhne et al., 2010), all of which showed the spelling tree through the visual domain. Owing to the intuitive mapping, this need was eliminated here.

With the use of the dynamic stopping method in Session 2, the ITR increased to an average of 5.26 bits/min and a maximum of 7.55 bits/min, which comes close to currently available covert visual BCI systems (Liu et al., 2011; Treder et al., 2011).

Front-back confusion

Learning from Schreuder et al. (2010, 2009), the stimuli were optimized for better performance. As shown in Figure 7, the number of detrimental factors was limited. However, speakers with the same angle from the frontal and posterior mid-line – speakers 1 and 3 or speakers 4 and 6 in Figure 1 – were confused relatively often. This type of error is well known in hearing research as the front-back confusion and results from an ambiguity in primary interaural differences, ITD in particular (Wightman and Kistler, 1999).

It was shown that differentiation in the front-back plane relies on at least two features: (1) spectral cues in the range from 8 to 16 kHz, referred to as the extended high frequency range (EHF) (Langendijk and Bronkhorst, 2002) and (2) small head movements (Wightman and Kistler, 1999). Bearing in mind the potential end-user and the short stimulus length, the latter may be out of reach.

However, for the first feature the noise component of the stimuli used in the current study could be considered sub-optimal. Though with aging the sensitivity to this EHF is reduced for a considerable number of people (Fozard and Gordon-Salant, 2001), the upper bound on the noise filter should be increased to 20 kHz to provide a more optimal range.

Furthermore, it is known that for posterior sound sources the intensity of high frequencies is reduced when compared to frontal sources (Blauert, 1983). In this study, the frontal speakers (1 and 6) both had lower frequency ranges in the stimulus complex than their respective confusers in the back (speakers 3 and 4), which may also account for some of the front-back confusion.

As shown in (Carlile et al., 1997), most front-back errors occur for locations within 30° of the median plane. Unfortunately, the before mentioned speakers in this study fall exactly within this range, which introduces a tertiary source for the found confusion.

A simple remedy could be to restrict all speakers to the frontal half, as done offline before (Schreuder et al., 2010; Belitski et al., 2011). When strictly adhering to the azimuth plane, this leads to a decreased angle between neighboring speakers. For a large number of classes it would be an interesting extension to use the vertical plane as an additional dimension to sustain the number of classes without compromising the spatial resolution of human hearing (Grantham et al., 2003). This is however not without caveats itself, as up-down localization may suffer from similar problems as front-back localization.

Dynamic stopping method

Increase in performance due to any dynamic stopping method is promising but largely neglected in literature. Though several methods have been introduced before (Serby et al., 2005; Lenhardt et al., 2008; Zhang et al., 2008; Liu et al., 2010), they seldom find their way into online studies other than for the purpose of validating the methods themselves. With an early decision in 41% of all trials and a false positive rate of 5%, we succeeded in constructing a conservative yet effective method for dynamic stopping. The method is loosely based on Lenhardt et al. (2008). However, instead of one global threshold, subject- and iteration specific thresholds are used. By taking a relative confidence measure (p), the method is robust against

biases. The results show again that dynamic stopping can significantly increase performance.

Performance in traditional visual ERP application is generally good even at a low number of iterations. However, things get more difficult when overt attention is impossible (Treder and Blankertz, 2010; Brunner et al., 2010), where performance breaks down and a quality check as provided by a dynamic stopping method may be desirable. The same can be said for auditory paradigms, where covert attention is the standard and the resulting ERPs are less strong. Dynamic stopping methods should thus find their way into BCI research and applications.

Here a decision is enforced after 15 iterations, but an additional benefit of dynamic stopping methods could be to refrain from any decision if no threshold is surpassed. It might prove useful to allow for trials where no decisions is made, to implement a no-control state or to simply refrain from taking uncertain decisions. This could be particularly useful in long-term use of a system, where at times the user may be in a no-control state (Huggins et al., 2011).

Towards a purely auditory modality

Most auditory BCI spellers described in literature depend to a certain extent on the visual modality for presenting a helping matrix (Belitski et al., 2011; Furdea et al., 2009; Höhne et al., 2010; Klobassa et al., 2009; Kübler et al., 2009) to the user. By designing an intuitive interface, presentations of such a (static) matrix was not necessary in the current study. Users could easily learn the required steps for writing a letter. Sentence 1 and 2b were written without any visual feedback other than the writing progress. Although this requires some additional time to present to the user his current position in the spelling tree, it represents a new step towards a purely auditory BCI. Though not presented here, it is worth noting that the system has been used successfully by a blind collaborator to spell a sentence, using exclusively the auditory cues.

Of course, a completely auditory BCI is not always required. Residual eye control may not suffice for a visual BCI, but could be good enough for presenting a static matrix. Furthermore, there are cases imaginable where auditory stimulation may be preferable over visual ones even when the visual modality is in tact. For instance, auditory stimuli may be easier to ignore and less intrusive when the user is engaged in a (primary) visual task such as reading.

In a recent study it was shown that auditory distraction has no effect on BCI systems based on mental tasks Friedrich et al. (2011). It is not clear if this holds true for auditory ERP BCI systems, as interaction in the same modality occurs. The level of distraction could depend on the relative loudness and the user's ability to concentrate. That this is not necessarily a problem became obvious when subject VPkw helped us out in a demonstration for the local television. She performed a near-perfect spelling session, whilst TV cameras were being set up in the room and the crew was talking.

Though not explicitly tested here, the current setup with added pitch cues may be particularly suited for patients that are practiced musicians, as their improved tone hearing can increase the benefit from these pitch cues (Micheyl et al., 2006).

Future recommendations

Although the AMUSE paradigm was now embedded in a speller system, it may prove useful in other domains. For instance, the speaker distribution on the azimuth plane could provide a user with an intuitive set of commands for spatial navigation. Also, by including secondary ERP signals such as the ErrP (Schmidt et al., 2011) for post-hoc analysis of a decision, the system could be expected to increase in usability.

The SOA of 175 ms used in this study was not optimized. Any SOA represents a trade-off between single-subtrial SNR and the number of repetitions needed. On the one hand, each subtrial is expected to contribute more class-discriminant information with a longer SOA. This can be explained by changed average amplitudes of target components, e.g. increased target P3 and N2, due to larger target-to-target distances (Gonsalvez and Polich, 2002; Allison and Pineda, 2006) and longer inter-stimulus intervals (Polich et al., 1991). But even other non-target ERPs can have an influence on the classification accuracy, e.g. the sensitivity of N1 and P2 to SOA and target-to-target distances (Budd et al., 1998; Allison and Pineda, 2006). On the other hand, the BCI system agglomerates evidence over time by repeated presentation of the stimuli. With a fast SOA, more repetitions can be performed within the same time, which compensates for the lower information content in each individual subtrial. Due to increased overlap in fast paradigms, this interaction has to

be considered carefully. As the best SOA value is probably subject dependent, an individualization for patients or other long-term users seems advisable.

Investigations of the basic ERP characteristics of ALS patients report a decrease of amplitude not only for the P3(a) (Silvoni et al., 2009) component, but also for the N1 and MMN (N2) components (Raggi et al., 2008) in auditory and visual oddball paradigms. An increase of P3 latency has also been described (Hanagasi et al., 2002; Paulus et al., 2002), even in direct correlation with ALS severity and months from disease onset (Raggi et al., 2008). Still, reports on ERP based BCI systems for patients are encouraging. Silvoni et al. (2009) found no correlation between clinical data and BCI performance for 21 early and mid stage ALS patients in an auditory ERP BCI, even though the P3 amplitude was reduced when compared to healthy controls. Furthermore, several applications for ALS patients have been published that show the feasibility of visual and auditory ERP paradigms (Nijboer et al., 2008; Sellers and Donchin, 2006). The question if ERP paradigms are suited for late stage ALS patients still remains open.

Recently an alternative auditory method was introduced, called PASS2D (Höhne et al., 2011), which is based on (Schreuder et al., 2010). Instead of using a free-field setup, a headphone setup was tested using cues differing in two dimensions (tone and location). The PASS2D and AMUSE paradigm have similarly high performances, which are competitive with state-of-the-art visual, covert attention BCIs. They furthermore complement each other in several interesting aspects. A headphone setup is portable, small and requires very little alteration to the home environment. It may be better suited for mobile user. On the downside, headphones may lead to social exclusion as the user might have limited perception of environmental sounds. The AMUSE paradigm is less mobile and requires an initial mechanical setup to position the speakers around the user. However, no social exclusion occurs, as the user can perceive stimuli and environmental sound at the same time. As end-users might be relatively immobile, the choice between AMUSE and PASS2D depends on the user's liking.

In conclusion, the AMUSE paradigm complements the current visual BCI systems with a realistic and high performing alternative. The majority of users can successfully control the system and concrete steps of improve-

ment have been offered. Furthermore, a method for dynamic trial stopping was introduced, which significantly increased performance.

Acknowledgment

The authors would like to thank Thomas Denck, David List and Larissa Queda for their help with performing the experiments. Furthermore, they thank Klaus-Robert Müller and Benjamin Blankertz for their fruitful discussions on the methodology. This work was partly supported by the European Information and Communication Technologies (ICT) Programme Project FP7-224631 and 216886, by grants of the Deutsche Forschungsgemeinschaft (DFG) (MU 987/3-2) and Bundesministerium für Bildung und Forschung (BMBF) (FKZ 01IB001A, 01GQ0850) and by the FP7-ICT Programme of the European Community, under the PASCAL2 Network of Excellence, ICT-216886. This publication only reflects the authors' views. Funding agencies are not liable for any use that may be made of the information contained herein.

References

- Acqualagna, L., M. S. Treder, M. Schreuder, and B. Blankertz (2010). A novel brain-computer interface based on the rapid serial visual presentation paradigm. In *Conf Proc IEEE Eng Med Biol Soc*, Volume 1, pp. 2686–2689.
- Allison, B. Z. and J. A. Pineda (2006). Effects of SOA and flash pattern manipulations on ERPs, performance, and preference: implications for a BCI system. *Int J Psychophysiol* 59(2), 127–140.
- Belitski, A., J. Farquhar, and P. Desain (2011). P300 audio-visual speller. *J Neural Eng* 8(2), 025022.
- Blankertz, B., G. Dornhege, M. Krauledat, M. Schröder, J. Williamson, R. Murray-Smith, and K.-R. Müller (2006). The Berlin Brain-Computer Interface presents the novel mental typewriter Hex-o-Spell. In *Proceedings of the 3rd International Brain-Computer Interface Workshop and Training Course 2006*, pp. 108–109. Verlag der Technischen Universität Graz.
- Blankertz, B., S. Lemm, M. S. Treder, S. Haufe, and K.-R. Müller (2011). Single-trial analysis and classification of ERP components – a tutorial. *Neuroimage* 56, 814–825.
- Blankertz, B., M. Tangermann, C. Vidaurre, S. Fazli, C. Sannelli, S. Haufe, C. Maeder, L. E. Ramsey, I. Sturm, G. Curio, and K.-R. Müller (2010). The Berlin Brain-Computer Interface: Non-medical uses of BCI technology. *Front Neuroscience* 4, 198.

- Blauert, J. (1983). *Spatial Hearing: The Psychophysics of Human Sound Localization*. Cambridge, MA, US: MIT Press.
- Brainard, D. H. (1997). The psychophysics toolbox. *Spat Vis* 10(4), 433–436.
- Brouwer, A.-M. and J. B. F. van Erp (2010). A tactile P300 brain-computer interface. *Front Neuroscience* 4(19), 036003.
- Brunner, P., S. Joshi, S. Briskin, J. R. Wolpaw, H. Bischof, and G. Schalk (2010). Does the "P300" speller depend on eye gaze? *J Neural Eng* 7, 056013.
- Budd, T. W., R. J. Barry, E. Gordon, C. Rennie, and P. T. Michie (1998). Decrement of the N1 auditory event-related potential with stimulus repetition: habituation vs. refractoriness. *Int J Psychophysiol* 31(1), 51 – 68.
- Carlile, S., P. Leong, and S. Hyams (1997). The nature and distribution of errors in sound localization by human listeners. *Hear Res* 114(1-2), 179 – 196.
- Cedarbaum, J. M., N. Stambler, E. Malta, C. Fuller, D. Hilt, B. Thurmond, and A. Nakanishi (1999). The ALSFRS-R: a revised als functional rating scale that incorporates assessments of respiratory function. *J Neurol Sci* 169(1-2), 13 – 21.
- Cincotti, F., L. Kauhanen, F. Aloise, T. Palomäki, N. Caporusso, P. Jylänki, D. Mattia, F. Babiloni, G. Vanacker, M. Nuttin, M. G. Marciani, and J. del R. Millán (2007). Vibrotactile feedback for brain-computer interface operation. *Intell. Neuroscience* 2007, 12 pages.
- Farwell, L. and E. Donchin (1988). Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalogr Clin Neurophysiol* 70, 510–523.
- Fozard, J. and S. Gordon-Salant (2001). *Handbook of the psychology of aging* (5th ed.), Chapter Changes in vision and hearing with aging, pp. 241–266. San Diego, CA, US: Academic Press.
- Friedrich, E. V., R. Scherer, K. Sonleitner, and C. Neuper (2011). Impact of auditory distraction on user performance in a brain-computer interface driven by different mental tasks. *Clin Neurophysiol In Press, Corrected Proof*.
- Furdea, A., S. Halder, D. J. Krusienski, D. Bross, F. Nijboer, N. Birbaumer, and A. Kübler (2009). An auditory oddball (P300) spelling system for brain-computer interfaces. *Psychophysiology* 46, 617–625.
- Gonsalvez, C. and J. Polich (2002). P300 amplitude is determined by target-to-target interval. *Psychophysiology* 39(03), 388–396.
- Gonsalvez, C. J., R. J. Barry, J. a. Rushby, and J. Polich (2007). Target-to-target interval, intensity, and P300 from an auditory single-stimulus task. *"Psychophysiology"* 44(2), 245–50.
- Grantham, D. W., B. W. Y. Hornsby, and E. A. Erpenbeck (2003). Auditory spatial resolution in horizontal, vertical, and diagonal planes. *J Acoust Soc Am* 114(2), 1009–1022.
- Guger, C., S. Daban, E. Sellers, C. Holzner, G. Krausz, R. Carabalona, F. Gramatica, and G. Edlinger (2009). How many people are able to control a P300-based brain-computer interface (BCI)? *Neurosci Lett* 462, 94–98.
- Guo, J., S. Gao, and B. Hong (2010). An auditory brain-computer interface using active mental response. *IEEE Trans Neural Syst Rehabil Eng* 18(3), 230 –235.
- Halder, S., M. Rea, R. Andreoni, F. Nijboer, E. M. Hammer, S. C. Kleih, N. Birbaumer, and A. Kübler (2010). An auditory oddball brain-computer interface for binary choices. *Clin Neurophysiol* 121(4), 516–523.
- Hanagasi, H. A., I. H. Gurvit, N. Ermutlu, G. Kaptanoglu, S. Karamursel, H. A. Idrisoglu, M. Emre, and T. Demiralp (2002). Cognitive impairment in amyotrophic lateral sclerosis: evidence from neuropsychological investigation and event-related potentials. *Brain Res Cogn Brain Res* 14, 234–244.
- Hill, N., T. Lal, K. Bierig, N. Birbaumer, and B. Schölkopf (2005). An auditory paradigm for brain-computer interfaces. In Y. W. Saul, L.K. and L. Bottou (Eds.), *Advances in Neural Information Processing Systems*, Volume 17, Cambridge, MA, USA, pp. 569–576. MIT Press.
- Hinterberger, T., N. Neumann, M. Pham, A. Kübler, A. Grether, N. Hofmayer, B. Wilhelm, H. Flor, and N. Birbaumer (2004). A multi-modal brain-based feedback and communication system. *Exp Brain Res* 154(4), 521–526.
- Höhne, J., M. Schreuder, B. Blankertz, and M. Tangermann (2010). Two-dimensional auditory P300 speller with predictive text system. In *Conf Proc IEEE Eng Med Biol Soc*, Volume 1, pp. 4185–4188.
- Höhne, J., M. Schreuder, B. Blankertz, and M. Tangermann (2011). A novel 9-class auditory erp paradigm driving a predictive text entry system. *Frontiers in Neuroscience* 5(0).
- Huggins, J. E., P. A. Wren, and K. L. Gruis (2011). What would brain-computer interface users want? opinions and priorities of potential users with amyotrophic lateral sclerosis. *Amyotroph Lateral Scler.* in press.
- Kanoh, S., K. Miyamoto, and T. Yoshinobu (2008). A brain-computer interface (BCI) system based on auditory stream segregation. In *Conf Proc IEEE Eng Med Biol Soc*, pp. 642–645. IEEE.
- Kim, D.-W., H.-J. Hwang, J.-H. Lim, Y.-H. Lee, K.-Y. Jung, and C.-H. Im (2011). Classification of selective attention to auditory stimuli: Toward vision-free brain-computer interfacing. *J Neurosci Methods* 197(1), 180 – 185.
- Klobassa, D. S., T. M. Vaughan, P. Brunner, N. E. Schwartz, J. R. Wolpaw, C. Neuper, and E. W. Sellers (2009). Toward a high-throughput auditory P300-based brain-computer interface. *Clin Neurophysiol* 120, 1252–1261.

- Krusienski, D. J., E. W. Sellers, F. Cabestaing, S. Bayouth, D. J. McFarland, T. M. Vaughan, and J. R. Wolpaw (2006). A comparison of classification techniques for the P300 speller. *J Neural Eng* 3(4), 299–305.
- Kübler, A., A. Furdea, S. Halder, E. M. Hammer, F. Nijboer, and B. Kotchoubey (2009). A brain-computer interface controlled auditory event-related potential (p300) spelling system for locked-in patients. *Ann N Y Acad Sci* 1157, 90–100.
- Langendijk, E. H. A. and A. W. Bronkhorst (2002). Contribution of spectral cues to human sound localization. *J Acoust Soc Am* 112(4), 1583–1596.
- Ledoit, O. and M. Wolf (2004). A well-conditioned estimator for large-dimensional covariance matrices. *J Multivar Anal* 88, 365–411.
- Lenhardt, A., M. Kaper, and H. Ritter (2008). An adaptive P300-based online brain-computer interface. *IEEE Trans Neural Syst Rehabil Eng* 16, 121–130.
- Liu, T., L. Goldberg, S. Gao, and B. Hong (2010). An online brain-computer interface using non-flashing visual evoked potentials. *J Neural Eng* 7(3), 036003.
- Liu, Y., Z. Zhou, and D. Hu (2011). Gaze independent brain-computer speller with covert visual search tasks. *Clin Neurophysiol* 122(6), 1127–1136.
- Micheyl, C., K. Delhommeau, X. Perrot, and A. J. Oxenham (2006). Influence of musical and psychoacoustical training on pitch discrimination. *Hear Res* 219(1-2), 36–47.
- Middlebrooks, J. C. and D. M. Green (1991). Sound localization by human listeners. *Annu Rev Psychol* 42(1), 135–159.
- Münssinger, J. I., S. Halder, S. C. Kleih, A. Furdea, V. Raco, A. Hösle, and A. Kübler (2010). Brain painting: first evaluation of a new brain-computer interface application with als patients and healthy volunteers. *Front Neuroscience* 4(182).
- Murguialday, A. R., J. Hill, M. Bensch, S. Martens, S. Halder, F. Nijboer, B. Schölkopf, N. Birbaumer, and A. Gharabaghi (2010). Transition from the locked in to the completely locked-in state: A physiological analysis. *Clin Neurophysiol*.
- Nijboer, F., A. Furdea, I. Gunst, J. Mellinger, D. McFarland, N. Birbaumer, and A. Kübler (2008). An auditory brain-computer interface (BCI). *J Neurosci Methods* 167, 43–50.
- Nijboer, F., E. W. Sellers, J. Mellinger, M. A. Jordan, T. Matuz, A. Furdea, S. Halder, U. Mochty, D. J. Krusienski, T. M. Vaughan, J. R. Wolpaw, N. Birbaumer, and A. Kübler (2008). A P300-based brain-computer interface for people with amyotrophic lateral sclerosis. *Clin Neurophysiol* 119, 1909–1916.
- Nijholt, A., D. P.-O. Bos, and B. Reuderink (2009). Turning shortcomings into challenges: Brain-computer interfaces for games. *Entertainment Computing* 1(2), 85–94.
- Paulus, K. S., I. Magnano, M. R. Piras, M. A. Solinas, G. Solinas, G. D. Sau, and I. Aiello (2002). Visual and auditory event-related potentials in sporadic amyotrophic lateral sclerosis. *Clin Neurophysiol* 113, 853–861.
- Polich, J., T. Brock, and M. W. Geisler (1991). P300 from auditory and somatosensory stimuli: probability and inter-stimulus interval. *Int J Psychophysiol* 11(2), 219–223.
- Raggi, A., M. Consonni, S. Iannaccone, D. Perani, M. Zamboni, B. Sferazza, and S. F. Cappa (2008). Auditory event-related potentials in non-demented patients with sporadic amyotrophic lateral sclerosis. *Clin Neurophysiol* 119(2), 342–350.
- Schlögl, A., J. Kronegg, J. Huggins, and S. G. Mason (2007). Evaluation Criteria for BCI Research. In G. Dornhege, J. del R. Millán, T. Hinterberger, D. McFarland, and K.-R. Müller (Eds.), *Towards Brain-Computer Interfacing*, pp. 297–312. Cambridge, MA: MIT press.
- Schmidt, N. M., B. Blankertz, and M. S. Treder (2011). Online detection of error-related potentials boosts the communication speed of brain-computer interfaces. submitted.
- Schreuder, M., B. Blankertz, and M. Tangermann (2010). A new auditory multi-class brain-computer interface paradigm: Spatial hearing as an informative cue. *PLoS ONE* 5(4), e9813.
- Schreuder, M., M. Tangermann, and B. Blankertz (2009). Initial results of a high-speed spatial auditory BCI. *Int J Bioelectromagnetism* 11(2), 105–109.
- Schröder, M. and J. Trouvain (2003). The german text-to-speech synthesis system mary: A tool for research, development and teaching. *Int J Speech Tech* 6, 365–377.
- Sellers, E. W. and E. Donchin (2006). A P300-based brain-computer interface: initial tests by ALS patients. *Clin Neurophysiol* 117, 538–548.
- Sellers, E. W., T. M. Vaughan, and J. R. Wolpaw (2010). A brain-computer interface for long-term independent home use. *Amyotroph Lateral Scler* 11(5), 449–455.
- Serby, H., E. Yom-Tov, and G. Inbar (2005). An improved P300-based brain-computer interface. *IEEE Trans Neural Syst Rehabil Eng* 13, 89–98.
- Silvoni, S., C. Volpato, M. Cavinato, M. Marchetti, K. Priftis, A. Merico, P. Tonin, K. Koutsikos, F. Beverina, and F. Piccione (2009). P300-based brain-computer interface communication: evaluation and follow-up in amyotrophic lateral sclerosis. *Front Neuroscience* 3(0).
- Tangermann, M., M. Krauledat, K. Grzeska, M. Sagebaum, B. Blankertz, C. Vidaurre, and K.-R. Müller (2009). Playing pinball with non-invasive BCI. In *Advances in Neural Information Processing Systems*, pp. 1641–1648. Cambridge, MA: MIT Press.
- Treder, M. S. and B. Blankertz (2010). (C)overt attention and visual speller design in an ERP-based brain-computer interface. *Behav Brain Funct* 6, 28.

- Treder, M. S., N. M. Schmidt, and B. Blankertz (2011). Gaze-independent brain-computer interfaces based on covert attention and feature attention. submitted.
- Vlek, R., R. Schaefer, C. Gielen, J. Farquhar, and P. Desain (2011). Sequenced subjective accents for brain-computer interfaces. *J Neural Eng* 8(3), 036002.
- Wightman, F. and D. Kistler (1999). Resolution of front-back ambiguity in spatial hearing by listener and source movement. *J Acoust Soc Am* 105(5), 2841.
- Zhang, H., C. Guan, and C. Wang (2008). Asynchronous P300-based brain-computer interfaces: a computational approach with statistical models. *IEEE Trans Biomed Eng* 55, 1754–1763.

Table 3: **Grand average confusion matrix.** Confusion matrix of classification results over all subjects. Sensitivity, positive predictive value (PPV) and the occurrence (N) of each target are given for convenience.

		Selected class						Sensitivity	N
		1	2	3	4	5	6		
Target class	1	765	9	72	16	27	23	83.9	912
	2	18	881	26	29	37	72	82.9	1063
	3	24	18	543	12	41	9	83.9	647
	4	24	26	15	629	54	54	78.4	802
	5	12	8	10	12	539	7	91.7	588
	6	44	84	66	71	59	773	70.5	1097
PPV		86.2	85.9	74.2	81.8	71.2	82.4		

(A) Top-down view of the circular arena (65 cm radius) with six speakers (1-6) and a 60° angular separation.

(B) Trial structure for the sentence "Franz jagt im Taxi quer durch Berlin". It shows two trials with word sequences and a pseudo-random sequence of 6 stimuli (subtrials) with a 40 ms duration and 135 ms SOA. A timeline at the bottom shows the stimulus duration and SOA.

Figure 2.EPS

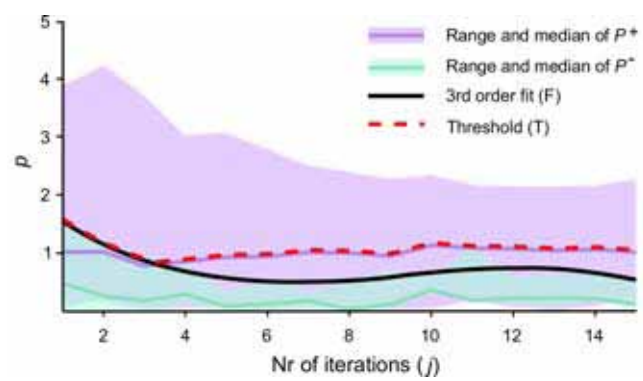


Figure 3.EPS

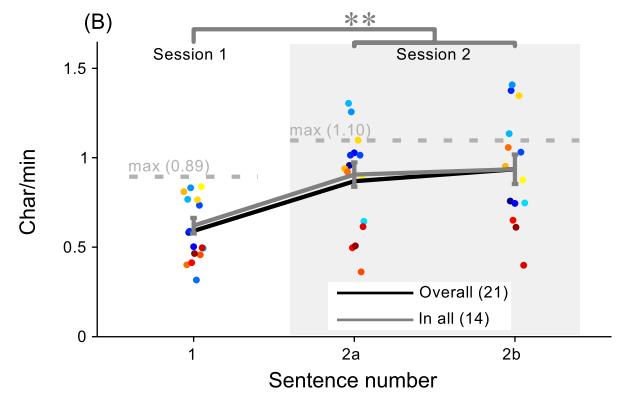
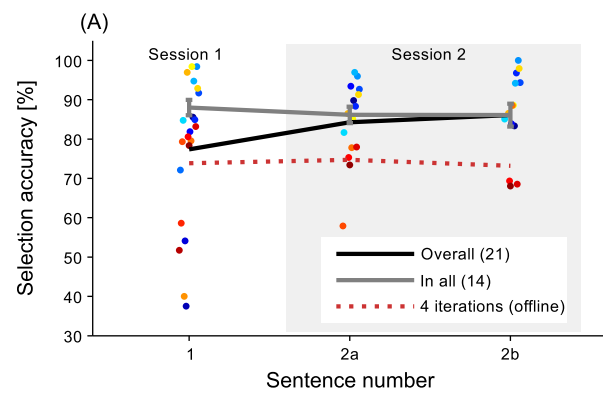


Figure 4.EPS

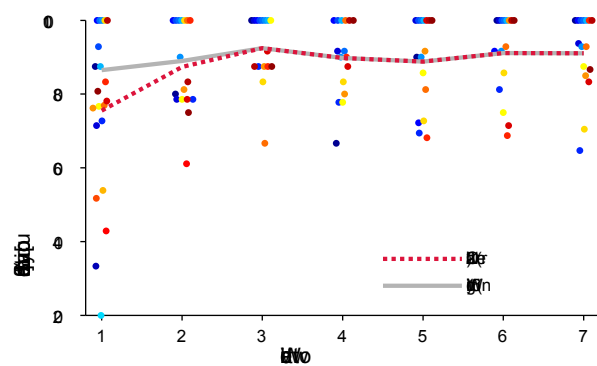


Figure 5.EPS

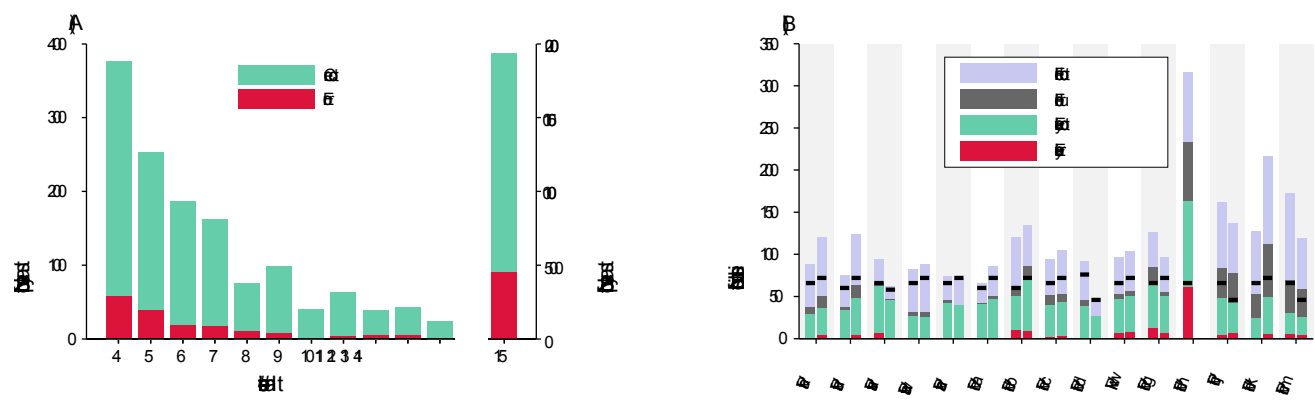


Figure 6.EPS

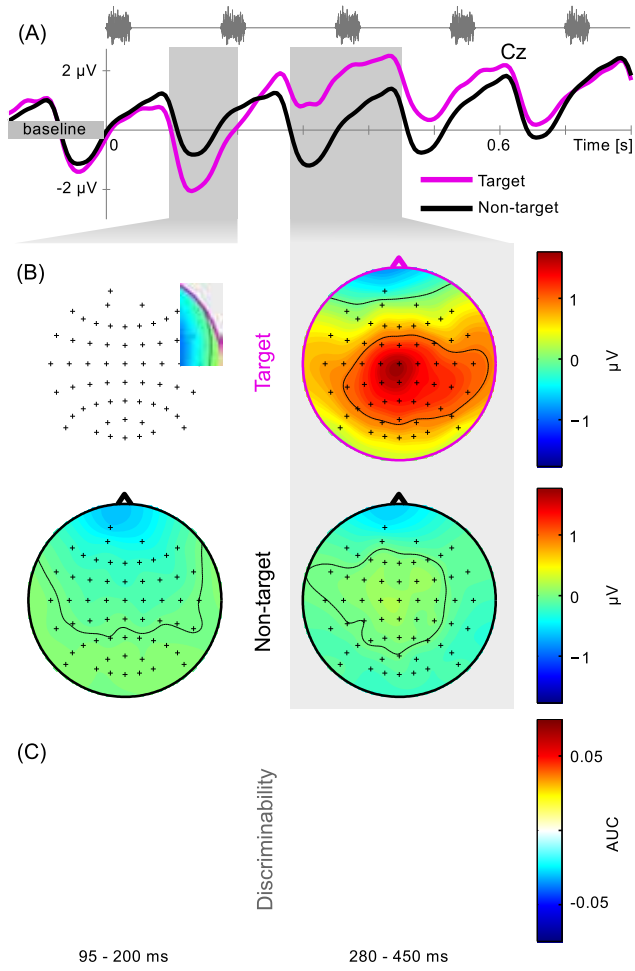


Figure 7.EPS

